DOCUMENT RESUME

ED 437 565 CE 079 653

AUTHOR	Border, Barbara
TITLE	The Status of Alternative Assessments through the 1990s: Performance and Authentic Assessments in Relation to Vocational-Technical Education Technical Skills, Workplace Skills, and Related Academic Skills.
INSTITUTION	Vocational Technical Education Consortium of States, Decatur, GA.
PUB DATE	1998-12-00
NOTE	53p.
AVAILABLE FROM	Vocational Technical Education Consortium of States, 1866 Southern Lane, Decatur, GA 30033-4097. Tel: 800-248-7701 (Toll Free).
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Education Work Relationship; *Educational Assessment; Employment Potential; Entry Workers; *Evaluation Methods; *Job Skills; Postsecondary Education; Secondary Education; *Standards; *Student Evaluation; Vocational Education
IDENTIFIERS	*Authentic Assessment

ABSTRACT

This study examines performance and authentic assessments related to job and academic skills. It begins with a general historical perspective that discusses the evolution of assessment systems used by education and industry in the United States, Germany, the United Kingdom, Australia, and Japan. The influence of federal policy also is discussed, and lessons from other countries are included. Finally, an in-depth report on authentic assessment and performance assessment today provides a view of trends in assessment, research results, and potential elements of assessment for consideration by states, the federal government, industry, and business for evaluation of both entry-level and advanced placement of workers. (Contains 78 references.) (KC)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 437 565

The Status of Alternative Assessments Through the 1990s:

Performance and Authentic Assessments in Relation to Vocational-Technical Education Technical Skills, Workplace Skills, and Related Academic Skills

December, 1998

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R McCage

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Vocational-Technical Education Consortium of States
Southern Association of Colleges and Schools
Decatur, GA

Barbara Border
Principal Investigator

CE 079653

BEST COPY AVAILABLE

The Status of Alternative Assessments Through the 1990s:

Performance and Authentic Assessments in Relation to Vocational-Technical Education Technical Skills, Workplace Skills, and Related Academic Skills

December, 1998

The Vocational-Technical Education Consortium of States
Southern Association of Colleges and Schools
Decatur, GA

Barbara Border
Principal Investigator

Acknowledgements

The following individuals have contributed to the development of this project:

Barbara Border, Principal Investigator

Ronald McCage, V-TECS Executive Director

Dr. McCage has had the foresight to move assessment from being an issue to an action item. He seeks to bring traditional written and performance/authentic assessments together to strengthen the data available for students, instructors, and employers.

Charles Losh, V-TECS Associate Director for Product Development

Dr. Losh has provided direction; he has contributed his knowledge related to crosswalking academics and the development of a New Scenario Assessment and Instructional Prototype for V-TECS.

Recognition of others:

Sri Ananda, WestEd

Dr. Ananda discussed the research results related to performance and authentic assessment projects WestEd conducted with California and the National Skill Standards Board.

Chris Olson

Dr. Olson reviewed the project and provided additional input.

Joan Wills

Ms. Wills provided materials and resources that contributed to the project.

Foreword

The Vocational-Technical Education Consortium of States (V-TECS) has been involved with the development of the components of standards for over twenty-five years. In the late 1980s and early 1990s, they provided the forum for assessment through conducting four national conferences for all interested parties. The primary focus at that time was on assessment related to competencies, duty-task lists, performance objectives, and standards. V-TECS sought methods of assessing these elements to assist the vocational-technical education community in measuring the progress of students in terms that business and industry would sanction and understand.

During the 1980s, V-TECS began developing and validating test item banks for the duty-task lists. During the 1990s V-TECS developed assessment components for two national standards projects, the Electronics Technician National Standards; and Heating, Air-Conditioning and Refrigeration Technician Standards. They worked with the development of assessment scenarios for the Indiana Manufacturing Linkage Project and the Family and Consumer Sciences National Standards Project. From this involvement, a strong rubric for valid scenarios has been developed. V-TECS is currently developing assessment components for the state of South Carolina with business and industry. They are also working with Cisco Corporation in developing workplace competencies and assessment components for the Cisco Networking Academics Curriculum.

Table of Contents

Introduction 5

I. General Status of Testing and Assessment in Academic Education 6
- Early Testing Processes in the United States
- Evolution of Educational Tests and Assessments
  - Norm-Referenced Tests
  - Criterion-Referenced Tests and Assessments
  - Competency Testing in Academic Education
  - Authentic and Performance Assessments
- Academic Testing and Assessment Lessons from England

II. Testing and Assessment Related to Workplace Skills and Competencies in Education . . 10
- Work-place Education in the United States
- State Policy Influences on Testing and Assessment
- Resources

III. Testing and Assessment by Business, Industry, and Labor 17
- Background on Testing and Certification Practices
- Patterns of Skills Testing and Certification by Business and Industry
- Apprenticeship Programs
- Trade Association Credentialling Practices

IV. Lessons from Other Countries Related to Testing and Assessment of Occupational Skills 21
- International Context
  - Germany
  - United Kingdom
  - Australia
  - Japan
- Summary of the Lessons from Other Countries

V. Alternative Assessment and Testing Modes 26
- Assessments That Support Change in Learning Styles in School and in the Workplace
  - In School
  - In the Workplace
- Research to Support the Development of Authentic and Performance Assessments
- Types of Authentic and Performance Assessments
  - Portfolios
  - Project Assessment
  - Scenarios
- Vocational-Technical Education Alternatives in Action
  - Vocational-Technical Education Student Organizations
  - System for Item Banks
  - The Work Keys System
  - The Testing Program

VI. Implications and Recommendations 43

References 46

Introduction

At present, it is difficult for business and industry in the United States to understand what a potential worker possesses in terms of a particular set of required skills. A school's normative test scores do not assist the employer in determining whether to hire the graduates of that educational institution. A diploma or degree does not assist the employer in determining what knowledge and skills the potential employee possesses. Nor do the individual norm-referenced scores assist the employer in hiring, as the criteria is unknown. The employer is interested in the skills and knowledge required for the work to be done now and in the future. According to Daniel Yankelovich in *The Forgotten Half Revisited*, in the chapter, "The Impact of Trends and Forces," recent survey data reflect that attitudes about educational accomplishments, preparation for the workforce, the demand for higher educational standards, the importance of teaching values, job skills and training are all going a negative direction for the Forgotten Half of students....Almost half of all Americans, almost two-thirds of all employers, and more than three-fourths of all professors do not believe that a high school diploma is a guarantee that a student has learned the basics..." (Halperin, 1998).

Likewise, parents and students are equally concerned about what students are achieving through education that will prepare them for the "real world" after high school or college. Parents and students have been baffled by norm reference scores, finding them difficult to relate to the students' progress in school. Students feels disassociated with the learning process in which they have little input or understanding of the relationship of achievement and assessment.

Education has been called to task in this country by all groups to reform and to provide more meaningful educational experiences and greater achievement levels on the part of all students. State Departments of Education are increasing their roles in assessment and examining various methods to provide tangible benchmarks for young people to use to judge their educational experiences. Educators are expanding their theories about instruction and assessments linkages and the roles of students and teachers in the process. In authentic and performance-based learning situations, the teacher becomes the facilitator and the students, the initiators of activities.

To focus on the problem of training a skilled workforce in America, The National Skill Standards Board (NSSB) is developing Voluntary Partnerships for 15 industry clusters. The partnership includes employers, unions, workers, community, government, education, and training representatives. These partnerships are responsible for developing standards, assessments, and quality assurance systems within their respective industry clusters. The United States is very late in becoming involved in such activities. European countries, Australia, Canada, and Japan have been involved in setting standards and assessments for decades.

The standards will provide a communication device that industry, education, training providers, community groups, labor and management, and workers and potential workers will use as a common language about skills and work.

The United States' system of standards includes an assessment component. Assessments must meet professional, legal, and technical standards; as well as be reliable, valid, and fair. NSSB is encouraging innovative methods of assessment such as simulations, performance measurement, multi-methods of assessment, and innovative methods of assessment delivery. (NSSB, Annual Report 1997-1998)

This study is an update to a document developed in 1994, related to assessment, NSSB projects, and federal and state education entities. A general historical perspective discusses the evolution of assessment systems used by education and industry in this country and others. The influence of federal policy also is discussed.

Finally, an in-depth report on authentic and performance assessment today provides a view of trends in assessment, research results, and potential elements of assessment for consideration by states, the federal government, industry, and business for evaluation of both entry-level and advanced placement of workers.

I. General Status of Testing and Assessment in Academic Education

Early Testing Processes in the United States

According to Webster's II New Riverside University Dictionary, the definition of testing is, "a series of questions and problems intended to measure the extent of knowledge, aptitudes, intelligence, and other mental traits." The definition has been expanded to mean, "to gather the clearest, most precise, consistent, and meaningful data possible to answer questions about student performance," according to Donald L. Hymes (1991). He suggests that assessment which provides a variety of methods and techniques to measure knowledge, skills, and other traits, should be an integral part of instruction, as opposed to being imposed from a separate source.

Robert Hayes also sees assessing as part of the learning and instructional process. "The outcomes should be established against which to measure the progress" (1984). Testing has been used since 1854, when the Boston School trustees determined that written tests would supplement oral examinations to determine the quality of instruction students were receiving (Hymes, 1991). Standardized testing for large groups of students was implemented in the early 20th Century. Intelligence Quotient tests were designed to track students. The Army also adopted these tests and used them in World War I.

During the 1960s, testing of achievement became more widely used than in the past due to the use of competencies as a means to set the criteria for scoring. In the 1970s and 1980s, accountability spurred growth in the use of standardized tests to document gains in overall

student achievement. In 1983, *A Nation At Risk: The Imperative for Education Reform* caused great concern about student achievement. For the first time, states became directly involved in the debates about the quality of schools and accurate data related to student achievement. By 1990, a large number of states developed assessment programs and mandated testing.

The debate about testing began to gain attention when the call for school reform began after *A Nation At Risk* was published. By the late 1980s, many educators, legislators, and others began to question the value of testing since improvement in education could not reliably be connected to the tests (Pipho, 1988).

At this point, schools were asked to be accountable and to improve program results. This began the move to establish valid and reliable tests to provide measures which schools could use for accountability and for comparison. From the early 1980s, norm-referenced tests, criterion referenced tests, competency tests, authentic assessments, and performance-based assessments have evolved.

Evolution of Educational Tests and Assessments

Norm-Referenced Tests. The tests commonly used in the early to mid 1980s were norm-referenced tests. Most of these were developed by private publishers and included such achievement tests as the California Achievement Test (CAT), Stanford Achievement Test (SAT), Iowa Test for Basic Skills (ITBS), and the Comprehensive Test for Basic Skills (CTBS). These achievement tests were developed to measure student knowledge gains in a particular subject matter. Local scores were measured against national norms. It is estimated that more than 100 million norm-referenced tests were administered annually at a national cost of $700 to $900 million each year (Hymes, Chafin, & Gonder, 1991).

The major advantages of norm-referenced testing were objectivity, cost effectiveness, and familiarity, according to Dr. Donald Hymes (Hymes, Chafin, & Gonder, 1991). The tests were machine-scored outside the education institution. The tests were cost-effective in that they cost pennies per student and gave the schools an inexpensive means of accountability. The tests were easily administered by instructors. The same procedures were repeated each year, thereby requiring little preparation on the part of teachers or administrators.

As the number of testing systems grew and local districts began to want additional measurements, the criterion-referenced tests became a supplement to the norm-referenced tests being used by districts.

Criterion-Referenced Tests and Assessments. Criterion-referenced tests are designed to measure whether a student has mastered specific content. They do not compare one student with another as norm-referenced tests do, but rather, measure the extent to which an individual student has grasped the subject.

Norm-referenced testing compares achievements of students, whereas criterion-referenced testing compares student achievement with specific criteria and objectives. As larger numbers of community representatives, business representatives, parents, and policy-makers began to question the value of achievement and norm-referenced tests, the criterion-referenced test became more relevant.

Value was seen in judging a student's success against the objectives of subject areas. For example, a license to drive is not given to a potential driver who cannot perform the task of driving without causing an accident, even though the individual may have made enough points on the test to score above the national norm on drivers tests. The criterion-referenced test measures an individual's knowledge and skills in relation to the set of tasks being learned.

The advantage of the criterion-referenced test is that it can identify the skills and competencies yet to be learned, thus providing a map of the future for a student's next learning experiences.

In the 1990s, criterion-referenced testing is being expanded to move beyond the paper-pencil, multiple-choice test process. Authentic and performance-based assessments currently are being explored by the academic education as well as the vocational-technical education, as a means to better measure each student's attainment of content.

The competency test has one meaning for academic education and yet another for vocational-technical education.

Competency Testing in Academic Education. Competency tests include standardized testing programs and are designed to determine if a student has attained a specified level of proficiency, and therefore can be promoted or graduated. These tests were developed by many states in the 1970s and 1980s in relation to state board of education requirements for various levels of education. The concern with these tests was that mastery of the test items became the maximum levels of attainment, as opposed to the minimum level of proficiency to be tested.

Authentic and Performance Assessments. Authentic and performance assessments, although not quite identical twins, are similar in nature. In the early 1990s a distinction was made between authentic and performance assessments. In authentic assessment a student often completed or demonstrated a desired behavior based on objectives selected by the student. The student determined the topic, the time allocated, the pacing, and the conditions under which the desired work was to be accomplished (Meyer, 1991). Today, the direction of the authentic assessment may be controlled by state or local requirements as is the case in Vermont and Kentucky. The student may have the opportunity to determine which work sample or activity to include, but the objectives to be demonstrated may be preselected. The trend appears to be in the direction of stated objectives to which students respond.

In performance assessment, a student completes or demonstrates a behavior an assessor desires to measure. The student performs a given set of tasks on a given topic, within a given amount of

time, and under given conditions (Meyer, 1991). Performance assessment evaluates the student response to being examined, whereas authentic assessment discusses the context in which the response is performed. Not all performance assessments are authentic, but all authentic assessments are performance tests.

At the current time, definitions are more closely related with less distinction made between the two. Performance assessment measures include a student's active generation of a response that is observable or that results in the development of a permanent product. Authentic assessment measurement relates the nature of the task and context to "real world" problems or issues. Students may select from the list or be given a topic. Both authentic and performance assessments are considered to be part of viable alternative methods for assessment (Elliott, 1995). Common features identified for both performance and authentic assessment include student construction, verbal or written response, and direct observation of the student.

Authentic assessments represent performance, evaluate against identified standards, and help students learn to evaluate themselves and to improve as they learn further knowledge and skills (Darling-Hammond, 1994). Performance and authentic assessments are viewed as nearly synonymous and as a means to assist students in applying knowledge and skills, problem-solving, and other process skills with contextual situations. The major impetus for the performance assessment movement has been the need to reconnect large-scale and classroom assessment to learning so that assessment affects learning and enhances instruction (Fuchs, 1995).

In the United States, academic education has explored various methods of testing and assessment and has moved away from pencil-and-paper tests toward practical assessments. Meanwhile, England, which has had a long history of performance assessment, is moving more toward written testing processes.

Academic Testing and Assessment Lessons from England

Although many educators in the United Kingdom think performance assessment benefits students, critics are challenging the rigor of evaluations and advocating dropping practical tasks from examinations. This move is based not solely on educational reasons, but also on the need to simplify the process and save money. National examinations are becoming very costly and driving the trend to put authentic forms of assessment into reverse.

England's traditional public examination and elite examination have remained in the old performance assessment format. In 1979, Her Majesty's Inspectorate of Schools, a group with high prestige and influence in education, recommended that more performance assessment be included in testing. It took eight years for a new examination to be developed as the General Certificate of Secondary Education (GCSE). Every content area includes some performance elements in the examination. The examinations, unlike norm-referenced testing in the United States, are graded by the pupil's teacher, with monitoring done by a governmental group. Some

of the concerns about the performance portion of the portfolio include the inability of students to gain significant assistance with their essays before they are included for assessment. It was noted by Michael Cohen and Lauren Resnick, New Standards Project (Resnick, 1989), that new types of learning are taking place as a result of the new performance examination. The 10th and 11th grades are student-centered and motivate young people to learn and foster a broad range of skills. Inclusion of a performance-based section in national tests is viewed positively in England.

The Thatcher government enacted the 1988 Education Reform Act which mandated the creation of the National Curriculum and an associated assessment system for all students enrolled in grades 2, 6, 9, and 11. By 1991, Standard Assessment Tasks were identified. However, these items were to be introduced by teachers with no standardized instructions. This leads to inconsistent implementation.

The lessons from England are that performance assessment can work and that both teachers and students want this type of examination. This type of assessment is demanding a new professionalism of teachers. Performance assessment can be successful only if teachers, parents, and politicians all agree to its importance and move toward development and implementation.

Unfortunately, the lessons from outside the classroom are that performance assessment can be costly in both financial and teacher time resources. In spite of all efforts to ensure rigor, many will not see it as such (Nuttall, 1992).

II. Testing and Assessment Related to Workplace Skills and Competencies in Education

Work-Place Education in the United States

Federal Policy Influences on Testing and Assessment for Work Place Education. The evolution of performance-based accountability for occupational education can be traced through federal acts from vocational-technical education and related bills for Manpower Training (MDTA), Comprehensive Education and Training Act (CETA), and Job Training and Partnership Act (JTPA), sponsored by the Department of Labor (Jennings, 1993).

The first act to have influence on accountability and assessment for occupational education was the Smith-Hughes Act of 1917. Less emphasis was on the quality of student gains and more emphasis was on instructor certification, classroom equipment, and instructional time. Smith-Hughes also initiated a system of "clear-cut management and accountability" (Cuban, 1989). With the requirement for state plans, long-term intervention of federal policy with modest financial support was established. The main criteria for evaluation was that programs were developed to meet the current and projected manpower needs and job opportunities. Less than three percent of the funds were set aside for program evaluation.

A similar 1963 act was considered the beginning of modern federal roles in vocational education. The Vocational Education Act of 1968 fueled tremendous expansion of vocational education enrollment and expenditures. Accountability requirements were greatly strengthened and greater emphasis was placed on evaluation. Federal support would be allowed only for programs which prepared students for employment, completion of programs, and meaningful occupational choices. However, the amendments did not address any formal outcomes-based evaluation (Wirt, 1994).

The 1976 amendments contained a separate section on evaluation. Within a five-year period, every local vocational education program had to be evaluated by the appropriate state agency. Every program had to impart how it provided entry-level skills, to what extent students completed programs and found employment in occupations related to their training. A new annual state accountability report had to summarize all evaluation findings and describe how these would be used to improve programs (Wirt, 1994).

The 1984 Carl D. Perkins Act marked the beginning of a completely new type of vocational-technical education legislation with federal prescription intensified. The U.S. House of Representatives attempted to require specific and outcome-based evaluation requirements for occupationally-specific programs. They made a significant effort to distinguish between expected outcomes and performance standards. The 1984 act required the establishment of technical committees comprised of business, industry, and labor representatives to define job competencies models. This was a foreshadowing of the Business and Industry National Skill Standards programs.

The Job Training Partnership Act of 1982 set a new standard for performance-based evaluation and continues to be an influence on human resource programs. The criteria for performance standards is very specific in various sections of the act. Some programs for adults require outcomes to include placement in unsubsidized jobs, increased earnings, reduced welfare dependency, and acquisition of employability and basic skills. Youth programs require successful attainment of employment competencies, secondary and post-secondary completion, enrollment in advanced training, or enlistment in military service.

Shifts in accountability from 1917 to the present have been related to the desire to keep federal-level influence on the program at a high level. Concerns about effectiveness of the program regarding its relevance to labor market needs prompted some of the accountability language which has become more highly defined over the years.

As vocational-technical education has matured, the definition of quality programs has changed to mean equity, modernization, and student impact on job placement and competency attainment. Formerly, quality involved external or maintenance factors of facilities, equipment, and professional staffing. Two of the newer assessment goals include dropout prevention and academic achievement.

Federal influence on accountability for state-level and local-level vocational-technical programs has moved into a stronger position with each act. State education personnel have reacted to this by placing an emphasis on student outcomes as opposed to process outcomes related to the program itself (Wirt, 1994).

State Policy Influences on Testing and Assessment

Statewide Academic Testing Programs. The Education Reform Movement during the late 1980's and early 90's has spawned many statewide performance assessment programs. Performance assessments offer benefits beyond those of traditional testing. They link assessment with instruction and can be used to improve learning and instruction that is currently in progress (McLaughlin, et al, 1995).

Examples of this type of performance-based program developed in the past five years include those established for Kentucky, Maryland, and Vermont. Currently, other states are moving in this direction away from norm-referenced large scale testing. In certain states the move to performance standards has been accompanied with high stakes consequences, thus bringing controversy to developing performance-based assessment programs.

Kentucky, as part of the Kentucky Education Reform Act, developed six performance goals for all students to attain upon graduation from Kentucky schools. These are being assessed through portfolios, performance events or alternative portfolios at 4th, 8th, and 12th grades. (McLaughlin, et al, 1995)

Maryland, through its Maryland School Performance Program (MSPP), developed comprehensive student outcomes related to reading and writing, mathematics, social studies, and science. Assessment processes include norm-referenced testing, criterion-referenced performance assessments, and criterion-referenced minimal competency tests (McLaughlin, 1995).

Vermont, through the Vermont Assessment Program, implemented the use of standardized assessments and portfolios for 4th and 8th grade students (McLaughlin, 1995).

During 1998, Arizona developed a statewide assessment for graduation called AIMS. This is a high stakes assessment to be administered on a large scale. Controversy surrounds the move to this criterion-referenced test system because linkages with instruction appear to be vague to instructional staff and community groups. The major concern appears to be that resources may not be available to provide the instruction to link with this performance-based assessment plan. This situation relates to one of the major concerns by theorists regarding performance assessments, that of cost not only for the testing but also for associated instruction.

State Vocational-Technical Education Assessment Systems. Although federal policies had an initial thrust toward the development of state performance standards, the states were given the responsibility for establishing testing and assessment measuring processes. The 1984 Vocational

Education Act and the 1990 Amendments brought the role of states to the forefront in establishing measurement criteria for vocational education programs and student outcomes.

The 1990 federal legislation required states to adopt outcome measures for learning and competency gains and at least one additional area--either competency attainment or employment outcomes. Most states have developed measures in four or five areas. These include gains in attainment of academic skills, occupationally specific skills, general employability skills, rates of program completion, rates of job placement, and status of employment or further education. These require some form of testing or measurement for all but three areas, according to the National Center for Research in Vocational Education (NCRVE)..

Two studies identify testing and assessment policies of states: *Education-Driven Skill Standards Systems in the United States* (Border, 1993), and *Testing and Assessment in Vocational Education* (Wirt, 1994). *Education-Driven Skill Standards Systems in the United States* provides insights about the status of criterion-referenced testing and authentic and performance-based testing related to the skill standards established by states since 1984. *Testing and Assessment in Vocational Education* provides insights into assessment methods used by states related to requirements of federal acts, especially the 1990 amendments. Wirt (1994) examined the role of competency testing that is used differently in vocational-technical education than in academic education. He found that vocational-technical and academic education had commonality in how they used certain types of assessment. These included written testing for reliable measurement and the broad-range methods of observation and demonstration for additional assessment of student accomplishments.

Written tests are keyed to specific items of knowledge and skills needed on the job. Other assessment methods included organized events or competitions with ratings, classroom and laboratory work with performance elements, rating scales, and rating procedures.

Portfolios are developed based on student attainment of written and performance-based competencies, recommendations by educators and employers, resumes, and school work.

The debate in academic education over norm-referenced versus performance and authentic testing is paralleled in vocational education. There is, however, little indication of reliance on norm-referenced and standardized tests in the vocational education field. Only recently have vocational educators used statewide large scale test results to determine academic gains of students. This is considered to be a short term remedy, but other viable indicators are being developed by groups such as the Vocational-Technical Education Consortium of States (V-TECS) and others.

Competency testing in vocational-technical education has roots in the competency-based movement; whereas, competency testing in academic education stems from the mental testing movement that began in psychology at the turn of the century.

Competency testing in vocational education is designed to measure whether a student has the skills needed to perform particular job functions, rather than a norm-referenced comparison with others nationwide. Vocational education tests often are written performance or hands-on performance and authentic assessments. Written performance tests are used to verify knowledge of the skills required to perform specified work functions. Hands-on performance assessments may be limited due to resources and materials available for testing students. However, this type of assessment is the preferred method of testing. Authentic assessments in the form of portfolios, projects, and scenarios are often used to augment written performance assessments. There is a relationship of identified job tasks, instructional content, and individual test items.

Traditions of competency-based assessment in vocational education are older than the competency-based movement in the United States. Specific lists of job competencies are used as the content and performance level criteria. Performance levels have various definitions and applications. The philosophy and methods are totally different from academic traditions.

Vocational education emphasizes the skills to be learned as the basis for instruction and assessment. For the most part, what has been tradition in vocational education testing is now being sought to replace traditional testing in academic education.

Wirt (1994) found four types of environments for testing and assessment by states. These include: 1) states that encourage written testing of occupational skills; 2) states which mandate assessment at the local level, but leave open the type of assessment to be used; 3) states that encourage assessment, but do not mandate it; and 4) states which have no specific policy or program of encouragement. Even though states may not have an assessment system, they still have performance standards and measures, as required by federal law.

Although assessment is a key component of occupational skill standards systems, less than one-half of the states had developed assessment systems at the state level by 1993. State officials indicate that local districts develop these with employers and local district input. Twenty-one states indicated that they did have test banks for use in developing criterion-referenced written tests. Oklahoma had the most extensive system. During the early 1990s, others developed test item banks at the state level including Florida, Michigan, North Carolina, Utah, Vermont, and Wisconsin. Minnesota and New York had local test bank items. New Hampshire and Oregon were developing banks. Florida developed more than 30 test item banks coded to statewide curriculum frameworks. Michigan had an extensive microcomputer-based test item bank. North Carolina also had a computerized test item bank (Border, 1993).

In this same study, states indicated that types of tests or assessments used to determine mastery of occupation skills included written, computerized, cognitive, simulation, situation, actual performance, and combinations of tests and assessments. Forty-one states used written cognitive tests. Simulation is used by twenty-six states. Thirty-eight states used actual performance tests. More than one-half of the states used a combination of testing processes to determine mastery.

By 1998, the number of states engaged in establishing a vocational-education testing system has increased by only a small percentage; however, many states are involved with all of education in Education Reform performance assessment processes.

During the mid-1980s, groups of states began considering the development of jointly established performance assessment processes. The Vocational-Technical Education Consortium of States (V-TECS) Board of Directors has been examining the potential of various systems and development processes for use by states. V-TECS began developing criterion-referenced test item banks in 1984. During the mid-1980s, V-TECS hosted four national conferences on assessment, bringing together the major researchers and administrators interested in developing assessment programs at the state and local levels of vocational-technical education. V-TECS began its work on item banks in relationship to the duty and task lists. States contributed items that were then validated by other states and groups. Industry was involved with development as well as the validation processes. By 1998, the item banks have been expanded and many banks revised, based on new duty and task lists and standards.

The National Occupational Testing Institute (NOCTI) has long been a vehicle for administering hands-on performance assessment processes. These are valid tests, but are limited to certain occupational clusters, due to cost and other factors.

Testing and assessment of skills is tied directly to instruction in most states. Although five years ago few states were moving to large scale testing in vocational-technical education, greater interest is building in 1998. However, little evidence indicates that one state or field of vocational-technical education is moving totally away from the classroom-laboratory-work environment directly connected to the learning situation as the place to continue to assess student progress. Certifications of skills practices usually are in the form of profiles; check sheets; documented observations; and certificates with lists of competencies attained, computer printouts of skills mastered, or similar records. Most authentic assessments and records of mastered competencies are provided by local secondary or post-secondary programs. The move to develop a valid and reliable testing system is only beginning to surface in this country. Interest is being generated as business and industry seek to raise the level of skills that entry-level, as well as advanced-entry, employees have in relation to specific sets of work functions. This is especially true with the industries related to information technology, telecommunications, computers, and advanced technologies. The major drawback for state development of a sophisticated performance assessment system relates to cost and other resources required to create such a system. The impetus in 1998, however, is that the new Carl D. Perkins Act has strengthened the need for performance assessment data.

Resources

Wirt (1994) found that some states have centers for criterion-referenced test bank development, such as Oklahoma with more than 100 job-specific areas included in its test bank. It has developed three types of tests: 1) tests measuring gains in specific occupational competencies

taught in specific courses, 2) tests measuring levels of competency achieved by students who complete a series of courses, and 3) profiles displaying student competencies for use in seeking employment.

In general, Wirt found that four types of resources are provided by states: 1) competency test items banks; 2) other forms of competency tests, such as those developed by Oklahoma; 3) profiles or portfolios; and 4) competency-based curriculum materials which frequently include testing materials (Wirt, 1994). Through multi-state consortium activities, V-TECS has two-thirds of its task lists linked to performance items. Item banks have been imported from states, combined and further refined. This provides a valid and reliable data base for written-performance assessments.

Skills used in assessments by states include broad technical skills, integrated academics, vocational skills, and generic workplace skills. The most interesting trend that Wirt (1994) identified is related to the measurement of integrated academics. States had to show the academic gains of students enrolled in vocational-technical education. With little time to develop valid and reliable tests, most states turned to the already in place standardized tests being used to project academic gains of students enrolled in vocational-technical education. By 1998, how to test academic gains of vocational education students is a concern. However, as the V-TECS Snyder Taxonomy for Academic Skills has gained widespread credibility, interest is arising for the development of related performance assessment components.

Practices in testing and assessment changed in vocational education. State officials indicated that performance standards adopted at the state level, as a result of the 1990 federal amendments, caused expansion of testing and assessment. Most expansion, however, was within existing components of testing and assessment rather than in the creation of new ones.

Under the new Carl D. Perkins Vocational and Technical Education Act of 1998, the performance assessment standards are more rigorous. After states have established baseline data in quantitative terms or percentages with the U. S. Department of Education, they are expected to set in motion a continuous improvement assessment process addressing four components:

1. Vocational and Technical Education Proficiencies and Academic Proficiencies,
2. Attainment of G.E.D. or postsecondary certificates,
3. Follow up of program graduates, and
4. Demonstration of improvement in performance of special populations against existing benchmarks.

Postsecondary vocational-technical education as a training provider for individuals who go through the One-Stop Career Centers, will also be part of a required cross-evaluation process with the criteria in the Workforce Development Act and the Adult Literacy Act. Incentive grants with few strings attached are being provided to states that meet the assessment criteria in each of the three acts. This could prompt new levels of interest in assessment systems by states.

Certain states have used not only their resources to implement assessment programs, but have gone to outside entities for portions of their system. Ohio revised their Competency Assessment Program to include three subtests: Work Keys System (ACT), revised competency tests for occupational-specific measurement, and employability skills. They also are shifting their state assessment policy to strongly encourage the usage of these by local programs. Texas developed a student portfolio assessment component to accompany profiles of student competencies. Also, it launched a performance-based assessment program for generic workplace competencies. South Carolina is expanding the number of occupational areas in its test bank. In 1998, South Carolina has contracted with V-TECS to develop both written and performance occupational assessments in seven areas. Iowa added requirements for assessing student competencies in local programs. West Virginia now also uses student profiles as part of the assessment system. Kansas added generic workplace skills to its profiles of student competencies.

III. Testing and Assessment by Business, Industry, and Labor

Background on Testing and Certification Practices

The growth of private organizations engaged in credentialling has expanded considerably since 1965. At that time, only 120 private organizations were providing such services (Jacobs, 1992). The numbers of credentialling organizations has grown to include trade associations, labor unions, and others to ensure that a quality standard is set for a specified type of work. Testing and assessment is conducted in several forms. The credential may be a journeyman card or a certificate with identified skills specifically identified and verified. Beyond this process, individual businesses and industries establish their own criteria for skills required for certain positions. Common patterns are found in the study, *Industry Driven Skill Standards Systems* (Wills, 1993).

Patterns of Skills Testing and Certification by Business and Industry

The industrial sector in the United States is more likely to hire a credentialed individual than is the agriculture, retailing, or the hospitality sector. The close association with labor unions may be the general rationale for this observation. The study points out that engineering technicians, professional secretaries, and a host of others representing a variety of occupations across industries have set standards and certification.

The manufacturing sector is often involved in apprenticeship programs and uses the criteria from these groups as the assessment and credentialling process.

The largest certification programs are directly related to occupations and industries where the threat to regulate the industry has provided the impetus. The health care industry and the real estate field are two examples of those which strive to regulate their own group.

Another common pattern of certification programs reflect specialties such as appraisers, financial, and insurance service certification programs. Many certification programs are geared to areas where large numbers of workers are not performing the same function, such as the construction industry. This provides the overall supervisor with the benefit of having skilled workers with certified competencies performing work which may not be the supervisor's specialty. There are a number of legal issues which are addressed with the various types and levels of certification.

The industry association may be the certification body for a group. Some of these areas may be very small including locksmiths or farriers. Others may be broad, such as the Associated General Contractors (AGC) which is large and spans a large number of construction occupations nationwide.

The Wills study [Wills (a), 1993] indicates that when a professional group within the field has strong credentialling practices, then the occupations related to the field are more likely to model this behavior. For example, the health care occupations model the nursing credentials after doctors' credentials.

The Wills study [Wills (a), 1993] also shows that certain patterns of practice are in evidence. Most of the programs offer only one or two recognitions. These programs may be linked to time in the workplace and have alternate paths that allow the candidate to use schooling in lieu of part of the time requirements for the workplace. Most certification groups have a written test, have a form of required recertification at a specified period of time, and provide "grand fathering" for those already in the occupation. Most groups have a core of knowledge which must be mastered before the minimum qualifications are met. Some groups have even developed ties to colleges which grant credit for their programs. A few have career paths established. This usually requires further testing to qualify for higher-level credentials.

Many of the trade associations develop certification programs through a means other than a skill or competency-based program. They accredit a school or program as well as the instructional staff. Individuals can gain credentials through successful completion of an accredited program. The accreditation process is used by many private and public industry associations in the United States today. The National Automotive Technicians Education Foundation (NATEF) certifies training programs for automotive technicians.

A primary problem in these programs is moving to competency-based from time-based programs. However, apprenticeship programs have always combined competency attainment with time-on-task as a basis for certification.

Apprenticeship Programs

Apprenticeship programs have been involved with the establishment of skills, testing and assessment, and the credentialling of workers since the early part of this century. The

competencies and skills are identified jointly by labor and employer programs. The tests are not only recognized in a local area, but nationwide. Often, the tests are designed by groups at a national level with variations allowed for locality differences. The tests are both written and performance in nature. Individuals take these tests to move into entry-level journeyman positions in industry. Advanced credentialling may be partially obtained through the union-sponsored programs, but usually emphasis is on the time spent on the job.

Trade Association Credentialling Practices

Many national trade associations have long-term practices of testing and credentialling. Several have credibility amongst not only the industry and business groups but also with vocational-technical education. The study on *Industry Driven Skill Standards Systems in the United States* (Wills (b), 1993) identifies many of these groups. The following are highlighted in this review:

1. Associated General Contractors of America (AGC)
2. International Masonry Institute (IMI)
3. National Joint Apprenticeship and Training Committee of the International Brotherhood of Electrical Workers (IBEW)
4. American Welding Society (AWS)
5. National Institute for Certification of Engineering Technologies (NICET)
6. National Institute for Automotive Service Excellence (ASE)
7. National Automotive Technicians Education Foundation (NATEF)

The testing and assessment processes used by the Associated General Contractors (AGC) include an accreditation of programs in eight areas. The program must have clear goals and objectives, recognition by an accredited institution, industry resources, administration, instructional materials, instructional staff, facilities, equipment and learning resources. All of the eight areas must meet the AGC criteria. The AGC rejected the performance process early in its development because of the estimated $200-300 cost per person to take the test.

The International Masonry Institute (IMI) has established 13 areas of certification. The training is time-based and available at specific Joint Apprenticeship Training Sites. The materials are competency-based and tests measure job-specific criteria. Open entry-exit programs are available. The assessment process is a performance-based as well as a criterion-referenced written examination.

The National Joint Apprenticeship and Training Committee of the International Brotherhood of Electrical Workers (IBEW) has established several levels of certification. At each level the student must pass a series of national benchmarked tests. The program requires both classroom work and extensive on-the-job work. There are 12 criterion-referenced tests, related to the content of the curriculum, which are updated each year and used for testing. Also, there are observation and portfolio requirements related to on-the-job learning.

The American Welding Society (AWS) began its program in 1976. They have standards for inspector at two levels, with a third tier being developed to meet the International Standards Organization (ISO) criteria. There are seven levels in which the welder can be certified. These also are in compliance with ISO criteria. Certifications are renewed every three years. However, individuals must provide documentation of continuous employment in the welding field to be considered for recertification. The AWS finds this keeps the welders current with new technologies. A registry of individuals of AWS certification is maintained (Wills (b), 1993).

The National Institute for Certification of Engineering Technologies (NICET) has two levels of technologist in every field. Most have four levels with progress from entry-level to proficiency-level, to experienced-level, to supervisory-level. Testing and Assessment is based on job task competencies, technician work experience, and other experience, including military training.

Competency-based tests include those related to construction materials testing, engineering models, fire protection, geotechnical, industrial instrumentation, land management and water control, mechanical, transportation, and underground utilities. Additional newer areas include electrical power, building construction, and additional sub-fields in mechanical, building construction, computer, telecommunications, hydro, and highway bridge design.

The assessment processes are written multiple choice exams, using an open book. The tests are criterion-referenced. Performance tests were deemed to be too expensive. Therefore, a verification process with the person's immediate supervisor is used to determine the ability to perform on the job. This group is a true third party evaluation group with no direct ties to any discipline oriented membership organization [Wills (b), 1993].

The National Institute for Automotive Service Excellence (ASE) and its partner the National Automotive Technicians Education Foundation (NATEF) are recognized for quality in the area of certification. NATEF is an independent organization supported by industry organizations.

The NATEF Board sets minimum standards for program certification, plan the self-evaluations of programs, train evaluation team leaders in each state, set standards for evaluation team leaders and members, develop methods of measuring evaluations, and set procedures for appeal of certification decisions. ASE focuses on technician certification, including eight automotive, six heavy-duty truck, two body-paint, and three engine machinist specialty areas. The tests are criterion- referenced multiple choice. They investigated the possibility of using performance assessments and found the cost to be prohibitive. Although the ASE tests may be taken at any time, a documented two-year work experience is required before certification can be received [Wills (b), 1993].

The National Institute for Metalworker Skills (NIMS) has established a very extensive effort to certify technicians in this field. They used the ASE model for establishing their organizational structure. A similar effort is underway in the Heating, Air-Conditioning and Refrigeration Industry.

IV. Lessons from Other Countries Related to Testing and Assessment of Occupational Skills

International Context

Several international organizations have been involved with the development of cross-national standards which directly impact the ability of an individual or firm to practice a trade, produce, or sell services in a country other than their own. The International Standards Organization (ISO) has been identified as an important certifying body. They have developed five international standards, including ISO 9000 to 9005. These provide guidance for development of an appropriate quality management program for a supplier's operation. The ISO criteria was originally developed in 1946 and was designed for use in two-party contractual situations and for internal auditing. ISO standards often are viewed today as the benchmark to strive toward by companies and workers in all countries (Wills and Sheets, 1993).

The American National Standards Institute (ANSI) and the American Society of Quality Control (ASQC) are taking primary responsibility for developing guidelines for using quality assurance principles in certification of education and training services.

Many studies have highlighted the skill standards and processes developed by Germany, England, and Denmark. The European system for educating individuals for skilled work has been heralded as the benchmark to be used by the United States. As in all types of systems, the processes are designed for the needs of their country, their situations, and the individuals involved. The following is an overview of systems and related assessments in Germany, the United Kingdom, Australia, and Japan (Wills and Sheets, 1993).

Germany. The German Duel System is broadly recognized for the apprenticeship systems it has established. This system dates to the middle ages, giving great strength to the institution. In the 1950s and 1960s the current system was developed to involve education, employers, unions, and government. The vocational training part of the duel system is supported with education legislation requiring state governments and federal-state cooperative arrangements. This structure for vocational education would be very much like that of the United States. The premise is to train German youth before they leave formal education. The teaching-training process involves two to three years of vocational training within private firms, business and labor organizations, and schools with related educational outcomes (Munch, 1991).

Each German state develops its own curricula, through centralized commissions of teachers or through special curriculum institutes. A degree of uniformity is provided as the Standing Conference of Ministers of Culture oversees the processes by states. All students must be able to pass qualifying examinations. There are contracts and a common framework for articulation between vocational school curricula and the training company.

Federal regulations require that competency bodies such as the Chambers establish a certification system for all apprentices within their regions. The regulations establish minimum requirements for admission to examinations, the composition of examination boards, and the structure and content of the examinations. Although the core content and structure of the examination is created by federal regulations, the local examining boards are responsible for actually developing and administering the examination. Most of the Chambers for each of the industries have created centralized committees to develop additional examinations for all occupations. Therefore, most of the exams given in localities are still developed by a central body for the entire country.

Examinations are given to apprentices before they are very far into the program to determine what progress has been made and what additional training is needed. Corrective action is then applied. In order to be admitted to the final examination, the apprentice must meet three major admission requirements: be near completion or at completion of training, have successfully completed the intermediate examination and required report book, and have legal registration of the training contract. The examination combines a written, criterion-referenced type test with multiple choice questions and short answers, and a performance component in the company training facility. Nearly 90 percent of the apprentices pass the final examination.

Three credentials are included in the portfolio of an apprenticeship completer: a school-leaving certificate, indicating the completion of mandatory school attendance; journey-level certification, indicating the successful completion of the examination and report book; and the employer recommendation, certifying performance (Wills and Sheets, 1993).

Credentials count in Germany. The lessons from this system include the concept that basic education levels are a prerequisite of training; employer participation and financial commitment is crucial; partnerships are successful; school-to-work transition must include investment in those being trained; and certification and examinations for excellence are mandatory (Nothdurft, 1989).

United Kingdom. The United Kingdom has established the National Council for Vocational Qualifications (NCVQ) to oversee the system of skill standards and credentialling. The current thrust is not only to examine very specific job titles for certification, but also, to begin establishing a new set of standards for a more generalized set of skills which could assist individuals who are moving across occupations. The National Vocational Qualification (NVQ) is used as a means to encourage individuals to be involved in continuous training, to expand knowledge and skills. In addition, the General National Vocational Qualification (GNVQ) is being established to test for practical competence and also academic ability. The English state that providing such a certificate will encourage more students to stay in school, further developing skills and acquiring knowledge (Wills & Sheets, 1993).

The major issue in the GNVQ examination is that one would be able to take this test and be qualified for postsecondary education just as one would be if taking the top comprehensive examination. At this point, there is a need to deal with the parity between the academic and

technical-vocational qualifying exams. Math appears to be the greatest concern, and several studies related to parity were being conducted in Scotland and England.

This system uses Industry Lead Bodies (ILB) to specify the standards for the NVQs for the occupations within their field. Lead Bodies are established to develop standards processes and the actual competencies to be measured for specific occupational clusters. The Lead Bodies have members representing industry, labor, and education (Peretz, 1992).

Award Bodies accredit the candidates. The Award Bodies develop the examinations, issue certificates of qualifications, and assure the validity and reliability of assessment instruments. The Award Body may be one of many groups which have been third party independent reviewers over a long period of time. These include the Royal Society of Arts; the Business, Education and Technical Council; and City and Guilds. City and Guilds award about 50 percent of the certificates.

Assessment is quite independent of training. A candidate is awarded credentials based alone on the outcome of the assessment. The assessment includes a work-site component paper and pencil test, and some form of performance assessment. These assessments include little criterion-referenced testing. Rather, the traditional practice has been to compare one person to another (Jessep, 1991).

The need to broaden the structure to encourage students to stay in school and to bring about a strengthened workforce is crucial to the economy of the country.

Australia. The Australian system is one which the United States modeled while designing our National Business and Industry Skills System. A minister of the National Training Board was "loaned" to the United States Department of Labor for a period of one and one-half years to assist in designing the system for the United States. Much of the language in federal legislation relating to development of the National Business and Industry Skills Board is comparable to that establishing the Australian board.

The Australian National Vocational Training System mandates the type of skills individuals in this system will possess when they complete a program. Career paths for workers are within specific industries and also may across industry lines. The system is a partnership between the National Vocational Training System, the National Training Board, Industries, and Labor.

The framework of the system includes setting competency standards with knowledge and skills, training and curriculum development, accreditation, training delivery, assessment, certification, and monitoring and review. Standards are set with each industry for all occupations within that industry. An individual moving toward certification as an electronics technician would determine in which industry to be certified. Training of the individual would be directed to the chosen industry, such as avionics or consumer electronics. The assessment would be in the industry, with a specialty in electronics technician. Australians state that this provides great

opportunity for developing a career path within an industry and to move to other areas within the same industry as positions become available (National Training Board-Australia, 1993).

Eight competency levels have been developed for the purposes of certification. Level 1 competencies mean that a person has an established work orientation and the knowledge and skills required to perform routine, predictable, repetitive, procedural tasks, theoretical knowledge and motor skills while working under close supervision. Level 3 requires that a person be capable of self-directed application including the selection and use of appropriate techniques and equipment required to perform tasks of some complexity and involving applied theoretical knowledge and motor skills. This worker must be competent, skilled, and autonomous. Usually a trade or journeyman certificate would be held by the worker. Level 8 competencies are professional qualifications and may include postdoctoral research, evidence of publications, and contribution to advancing knowledge in particular areas (National Training Board-Australia, 1993). The system is structured to include in the hierarchy, all workers--those with skills, as well as those who have college degrees.

Assessment for certification is competency-based, with a minimum of literacy requirements. The tests are more performance and less written in nature. The development of these assessments is a partnership effort involving vocational schools, industry, and labor unions. The Australians also have Competency Standards Bodies (CSB) which establish standards and testing processes to certify individuals. Members of CSB are appointed by the National Standards Board. The Bodies must meet a rigid criteria for the development of standards and assessments processes (National Training Board-Australia, 1993).

Japan. The Japanese have made a major commitment to upgrading and retraining employed workers, as well as moving new workers into the system. National educational standards and general academic qualifications and credentials play a central role in this process while vocational standards and qualifications play a minor role (Rosenbaum, 1991).

Standardized tests begin at the junior high school level and follow on through high school. Postsecondary education institutions, and even employers, use national standardized tests as the single most important passport for employment and further education and training (Resnick, 1989). The Ministry of Education has primary responsibility for developing and conducting these examinations with input from education, government, and industry associations. The emphasis is on general academics, only touching on broad vocational concepts. Ministry of Labor tests are given for trade skills resulting in certification. However, employers rarely view this certification as important (Wills & Sheets, 1991).

The Ministry of Labor offers examination based on previous education, training, and work experience. A specific length of employment is required before each trade test can be taken. If a worker has been enrolled in an accredited program, many times the individual would be exempted from written examinations (Wills & Sheets. 1993).

Summary of the Lessons from Other Countries

Germany, England, Australia, and Japan have been studied a great deal because of their prominence in the industrial world. Lessons from Germany identify strong ties between the assessment system and the apprenticeship program. The assessment instruments are used as both the credentialling tool and as an improvement tool as the student is learning. In Germany both the criterion-referenced test and the performance processes are valuable. Their system addresses very specific skills and specific academic knowledge. Rigor is expected and nearly all apprentices pass the final assessment.

The British have been restructuring their system and moving toward broad based skill standards to provide opportunity for students to continue schooling and choose from a greater variety of jobs. The United Kingdom is continuing its very specific system of the NVQs but is adding the GNVQs. The English use national normed tests, but also use performance tests.

The Australian system has undergone some changes since it made its start with a great deal of emphasis on the front part of the system, including the establishment of skill standards and competencies. They are moving ahead with both performance and related testing processes. Much of this is the responsibility of the Competency Bodies.

The Japanese system of rigor in the examination phase for all students and even those in employment, suggests that the most important part of the system is testing and assessment. The emphasis appears to be on standardized tests developed at the national level. Vocational education tests are established by the Ministry of Labor. These also are highly academic in nature with both academic and vocational portions of the test being paper and pencil tests.

General lessons from other parts of the world are that trade groups, such the European Community (EC), are busy examining how credentialling will allow for individuals across all their countries to be "portable." This has great interest for the United States as part of the North American Trade Treaty countries. The ISO criteria is becoming a very important standard for areas it addresses. Often, it is viewed as the worldwide benchmark for quality.

Although National Skill Standards are being established with coalitions of industry, education and training, labor, and others, it is difficult to design systems that are reliable and valid. United States industry and business are looking beyond the high school diploma and the college degree which do not tell the employer what the potential employee can contribute to their workplace. Likewise, students also want to know specifically how they are prepared for work. They want to understand the criteria against which they are being benchmarked; and they want to improve their skills to reach these standards, before exiting the schools and programs that prepare them for the real world. They increasingly want input regarding learning and self-evaluation methods to promote continuous improvement. Students desire to use contextual learning and apply facts to various situations and experiments, and to work with teams to accomplish learning. They are eager to take part in problem-solving activities, using knowledge and skills acquired as a basis

for learning. Industry and business in this country also want employees who have highly developed knowledge and skills related to their work, but also those who can apply this to various situations through various work functions. Thus problem-solving, creativity, team work, and continuous assessment for improvement are all part of the desired skills and knowledge for employees.

V. Alternative Assessment and Testing Modes

In partial answer to the education reform movement, industry's need for highly skilled employees working in a problem-solving team environment, and the demand by individuals to be part of the process of developing their own learning and evaluation environment, alternative assessment systems are being established. Alternative systems of assessment must relate directly to the learning process, must be clear and understandable, and must be measurable beyond attainment of factual knowledge. Also, they must have the rigor of validity and reliability. Researchers and educators see alternative assessments as key to improving learning for students, providing students and parents with more understandable data related to progress, and supplying richer data for employers to make decisions about employee hiring and placement.

Today, authentic and performance assessments are being drawn together as alternative assessment approaches. They include projects, portfolios, scenarios, observations (hands-on performance), and written performance (criterion-referenced) assessment methods. Emphasis is on what students can do, as well as what they know. These are excellent tools to use in providing meaningful information for students, teachers, parents, and others about what students have learned, and in relating to day-to-day instructional decisions being made about programs (Stiggens, 1993). Taylor (1994) suggests that authentic and performance tests must be so central to learning that the test is valued and used to further learning, as well as demonstrate it.

Performance and authentic assessments that model the real-world demand that there is integration of the application of knowledge and complex thinking, reasoning, problem-solving, and reflection skills. These assessments go beyond recall of facts, concepts, principles or procedures [WestEd (a), 1998].

Authentic assessments may be cumulative as in portfolios or projects; may require more complex answers; use authentic tasks; or mirror requirements for success at work or in the community, as in scenarios or observations at work sites. Performance assessments may be on-demand, reflecting cognitively complex situations, authentic and integrated, but to a lesser degree than cumulative assessments. Multiple choice tests can reflect authentic situations from real life that can be measured for relatively deep levels of understanding. These are written performance tests.

Students may also prepare a work sample which may be observed by the instructor, but may be self-evaluated by the student to increase self-reflection. Performance and authentic assessments

are flexible, allowing for a range of responses or performances within a range of demonstrable mastery [WestEd (a), 1998].

Assessments That Support Change in Learning Styles in School and in the Workplace

In School. Teachers can have a great deal of confidence in the changes they make in the instructional program with the aid of performance and authentic assessments. According to Fuchs (1995), teachers can make three types of decisions using authentic and performance assessments. These are instructional placement decisions, formative evaluation decisions, and diagnostic decisions. The assessments allow for teachers to understand what needs to be taught next. Teachers are able to monitor a student's learning, while instruction is underway, and can change the instruction program as needed. Specific difficulties can be pinpointed early and allow for remediation of the learning process. Fuchs provides three criteria that authentic and performance assessment need to meet to inform the instructional decision process. These include 1) measure important learning outcomes, 2) address all three purposes of assessment, and 3) provide clear descriptions of student performance.

Outcomes must be defined with assessment tasks matching relevant instructional goals. Performance standards need to be benchmarked to provide a criteria against which student progress can be compared. Decisions about students with disabilities may include determining whether their performance will be compared against a set benchmark or against their own progress. Scoring reliability also needs to be considered in making decisions in relation to assessment for students with disabilities (McLaughlin & Warren, 1995).

Currently, education is focused on performance and authentic assessment methods, with researchers working diligently to establish criteria for development of these assessments. Many theories have been brought forth for possible consideration. Action research studies completed by WestEd indicate that more than one type of assessment needs to be used to ensure validity and reliability in performance testing. Rubrics are strong for each type of authentic and performance assessment, but consideration must be given to the variables. The Linkages Project for Manufacturing administered through the State of Indiana has developed scenarios for assessment related to occupational skills. These efforts have moved forward the information base related to authentic and performance assessment and will inform the process for statewide and multi-state development of related assessments in the future.

Teachers, administrators and education policy makers embrace performance assessment because it allows for assessment in a quality manner of writing and communicating in English or in a second language; working cooperatively on a team; and completing laboratory work. Educators have become obsessed with performance assessment in the 1990s, much as they were with multiple-choice tests 60 years ago (Stiggins, 1995). Concern about the unreliability and

invalidity of carelessly developed subjective assessments has been issued by the assessment community, but for the most part is being unheard (Dunbar, Kortez, and Hoover, 1991).

A variety of methods indicating accountability, including teacher surveys, interviews, and extended case studies, pressure teachers and administrators to focus planning and instructional efforts on test content and preparing students to do well on the tests (Herman & Golan, 1991). Narrowing curriculum by over emphasizing basic-skill subjects and neglecting higher-order thinking skills is a practice used in many schools which serve high risk and disadvantaged students where there is the greatest pressure to improve test scores (Herman & Golan, 1991). When assessments model authentic skills, the content is learned in a context, the application teaches problem-solving, and the students model higher order thinking skills. Assessments which follow this type of model provide individuals with greater opportunities to exhibit their skills and knowledge in a larger context (Chapman, 1991).

In the Workplace. Likewise, the workplace is reflecting a new order of work. Functional work areas are developed and teams of workers with a variety of skills from research, operations, finance, and sales are placed together to organize and do required activities. They do not follow a structured set of detailed work orders from a supervisor, but consider the required outcomes, and design the solutions based on problem-solving, creativity, statistical projections, resources, and ability to achieve the outcomes. Work combines cognitive, technical, and process skills to determine what will be achieved, how it will be accomplished, and what the criteria for successful accomplishment includes.

As a result, simple skill performance or written performance assessments alone do not relate well to the multiple sets of skills required in work settings in the 1990s and in the 2000s. Self-assessment will continue to play key roles as individuals work more independently via telecommunicating experiences. Continuous improvement versus stagnation will be a key component related to independent self-evaluation. Alternative assessments will continue to play larger roles related to the diverse workforce needs.

Research to Support the Development of Authentic and Performance Assessments

Criticism of performance and alternative assessments relate to validity, reliability, and to scoring. Messick (1994) criticized performance assessment for the "task-specificity" involved that does not allow for a range of knowledge and skills to be demonstrated. Messick states that even though these assessments are authentic to the real world, they do not simulate all of the complexity of real world function. The concern is that all knowledge and skills are not tested. Messick also is concerned about the ability to use performance assessments across sites as well as tasks, bringing up the issue of reliability.

The question of capability versus capacity also arises. Capability refers to an individual's ability to successfully complete a task. Capacity does not refer to attaining specific skills, but the predictability as to the success in the future an individual may have in performing these and related tasks in a similar or a different environment.

Quality control lessons for performance standards have begun, but the rules are unclear for how to develop and use performance assessments (Messick, 1994). The need for these criteria has promoted national programs of research and development to refine the vision of the critical elements related to sound assessments, including performance and authentic assessments (Wiggins, 1993). Extensive work has been completed by WestEd on performance assessment. Their findings and established criteria will highlight part of this national effort toward improved validity and reliability for performance and authentic assessments. The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has also been examining the issues related to competencies and assessment related to the workplace. Katz and Chard have contributed to the dialogue, as have numerous other researchers who have been interested in authentic assessment processes for a number of years. All researchers have a common quest related to validity and reliability issues of performance and authentic assessments.

Quality assessments are built on current theories of learning and cognition and are grounded in views of what skills and capabilities individuals will need for future successes (Herman, 1992). Too many times assessments are judged by what they are not--standard, traditional multiple-choice. According to cognitive researchers, meaningful learning is reflective, constructive, and self-regulated (Herman, 1992). To know something is not just to have received information, but to have interpreted it and related it to other knowledge. Studies related to integration of learning and motivation highlight the importance of affective and metacognitive skills (Herman, 1992). Individuals who are not proficient at problem-solving and thinking may not be able to take cognitive information and use it in a metacognitive or affective manner. Competent thinkers and problem-solvers possess the disposition to use skills and strategies, as well as knowledge, in an affective manner. Groups may facilitate learning by modeling effective thinking strategies, layering complicated performances, providing mutual constructive feedback, and valuing the elements of critical thought (Resnick, 1989).

The question of validity relates to the appropriateness, meaningfulness, and usefulness of specific inferences. It is defined as, "The process of accumulating evidence to support inferences" (American Education Research Association, American Psychological Association, and National Council on Measurement in Education, 1985).

Validity is often posed as the big question by researchers in the area of alternative assessments. Can new assessment processes being developed accurately reflect the knowledge, skills, and abilities they are intended to measure? If assessments move away from the traditional processes of paper and pencil to authentic, performance, and group testing; will quality in measurement suffer? Some criteria has been established to judge the quality of the assessment strategy. This

criteria includes weighing consequences, fairness, and equity. Will social standards and equity issues overshadow how individuals think in complex situations and in problem-solving?

Content, predictive, and construct validity need to be considered in constructing authentic performance assessments (NSSAC and West Ed, 1998). Content validation relates to having adequate samples of the knowledge, skills, and abilities related to the field as part of the assessment (Wills, 1997). Predictive validity is particularly important for projecting future performance, using a current set of criteria to establish the assessment (Wills, 1997). Construct validity is unique in that it uses both analytic and quantitative data to determine the meaning of performance scores. Simply stated, the "construct," a psychological theory underlying the assessment instrument, is determined to be valid.

The other question is that of reliability. This element is often overlooked in examining performance and authentic assessments. Measurement errors that reduce reliability result from 1) inadequate or inappropriate selection of specific tasks for the knowledge, skills, and abilities base (content domain sampling), 2) disagreement among scorers (inter-rater reliability), or 3) differences in conditions under which assessments were conducted (standard operations) (NSSAC and WestEd, 1998).

Baker and Linn (1991) identified eight criteria to assist with determining validity and reliability related to assessments. These eight criteria are assessment consequences, fairness, content quality, content coverage, cognitive complexity, meaningfulness, transfer and generalizability, and cost and efficiency. These criteria provide a grounded set of elements by which to judge the points of reliability and validity in relation to performance and authentic assessments.

Yet another set of criteria to address validity of performance assessments has been developed by Baker & O'Neil (1993). Five characteristics are identified for valid performance assessments:

1. Have meaning and motivate high performance,
2. Require demonstration of complex cognition, applicable to important problem areas,
3. Exemplify current standards of content or subject matter quality,
4. Minimize facts of ancillary skills and irrelevant materials to focus the assessment.
5. Possess explicit standards for rating and judgment.

The scoring process is an important element of the performance and authentic assessment system. Scoring is considered a central element to the validity question. When large scale or high stakes assessments are involved, two methods may be used to assist in determining scoring validity. These include assembling a content expert panel to review scoring scales in relation to content being assessed. How well the scale can be used to score the level of difficulty is a central question to validity. The second method is to check for distribution of scores for implications of inequities in attention to gender, ethnicity, and difficulty levels (NSSAC and WestEd, 1998).

The sample of content being tested must reflect the total knowledge, skills, and ability base. Scorers need to be trained in the elements of scoring related to performance assessments to bring

about like scores for performance difficulty level assigned by independent raters. Conditions for testing in different sites must be monitored for equity (NSSAC and WestEd, 1998).

Transfer and generalizability must support other measurements used to determine an individual's abilities to apply knowledge and skills in various contexts. Cognitive complexity may be difficult to determine in assessing higher-level thinking. Content quality must reflect the best current understanding of the field. Content coverage must be considered. Using alternative assessments cannot leave large areas of content unattended. The contextualized assessments must be meaningful in education experience and motivational for performance. Finally, these assessments must be cost-effective and efficient. Performance and authentic testing is often identified as costly in time and materials. Will these issues overshadow results (Herman, 1992)?

Performance assessments must provide vision to facilitate improvement in instruction and learning. They must be designed to involve the student in problem-solving and in performance of tasks which are important to the student and perhaps to others (Baker, Dunbar, and Linn, 1991). Performance assessments are not all alike. They may not all have equal value to the curriculum or to the program, but have value to learning as well as to assessment. A criteria for evaluating performance assessments was developed by Baker, Dunbar, & Linn (1991). They have identified three broad areas to be considered in examining performance assessments: consequences, generalizability and transfer, and fairness of the assessment.

Consequences address the intended and unintended ramifications of measurement. The assessment must have promise for assisting with student learning. Face validity does not increase the intended effects of the assessment. Generalizability and transfer address the need for performance assessments to be applicable in several areas as there can be only so many performance assessment problems administered. Does the problem being solved in performance assessment have transfer to another skill? Fairness is the criteria applied to ensure that no minority, gender, or other bias is contained in the assessment. Practicality of administration of the assessment is yet another criteria to be considered. Will the student have time to complete this assessment within the school or work setting? Will the student have adequate resources to complete the problem? (Linn, 1991)

Richard J. Stiggins, of Northwest Regional Education Laboratory, advocates the performance assessment technique and has defined four components for consideration: a reason for the assessment, a particular performance to be evaluated, exercises that elicit that performance, and systematic rating procedures. Stiggins states that performance should be observed and graded on the spot, or recorded on video or audio tape for later scoring. Assessment or test administrators must be specially trained to administer such measurements (Hymes, 1991).

Edward D. Roeber, Michigan Assessment Program, sees many applications for the performance assessment. He identifies for career development such performances as applying for a job, interviewing for a job, and participating in a work team in both lead and follow positions on the team. Those who support the alternative assessment process see two major challenges:

1. The design of performance assessments that reflect the full domain of the curriculum and assistance with instructional strategies for teachers to use with individual students, and
2. The development of a scoring method for the performances to meet the demand for accountability; and to give a valid report of the achievement of each student, school, and school district. The performance assessment must almost meet the norm-referenced criteria to be a valid testing process (Roeber, 1995).

Grant Wiggins, National Center on Education and the Economy, states that authentic assessments must be judged upon their ability to closely match the standards and challenges of real life. The problem to be solved must replicate the challenges and standards of performance that typically face business people, scientists, community leaders, and others to have an authentic assessment. The authentic assessment may involve essays, reports, interviews, proposals, portfolios, and other supportive evidences. Wiggins (1989) states, "Evaluation is most accurate and equitable when it entails human judgment and dialogue, so that the person tested can ask for clarification of questions and explain his or her answers."

Stanley Rabinowitz, WestEd, concluded from the C-TAP project, that authentic and performance assessment must be system driven. He states that the system can be designed for increased reliability and validity through the triangulation approach, allowing for assessment of skills through multiple and complementary means. Through the C-TAP project, it was determined that one form of authentic assessment may not be comprehensive enough to cover all aspects of student learning (O'Neil, 1997).

Types of Authentic and Performance Assessments

Generally, authentic assessments require students to demonstrate what they can do as workers in real, out-of-school settings. Assessments may be so embodied in the instruction that they are nearly indistinguishable from it (Darling-Hammond, 1993). Students are involved in assisting with the selection of the topics and activities related to the assessment. The types of assessments may include portfolios, projects, scenarios, skill and workforce performance, diaries and presentations. The first three types identified hold the most interest at the present and are discussed below.

Portfolios. This type of authentic assessment is a purposeful collection of student work that shows the achievement or growth of a student. Portfolios provide student self-assessment and control of learning, support student-parent conferences, select students for special programs, certify students' competencies, demonstrate skills and abilities to employers, build student confidence, and evaluate progress (Arter, 1995).

Portfolios include a variety of student work related to multiple standards and even multiple topics on themes. Portfolios can provide a comprehensive view of a student's knowledge and skills related to standards that other types of authentic assessment fail to provide (WestEd (b), 1998).

According to WestEd, portfolios often include work product samples, writing samples, and career-related materials. Key features of an effective portfolio as established by WestEd include the following:

1. A variety of student work should be included in a portfolio to reflect multiple standards.
2. The portfolio should grow out of regular classroom work and work experiences.
3. Include documentation of performance improvement over time or overall achievement.

Portfolios may be treated as formative, as well as summative assessment material. They are viewed, critiqued, and input is provided to the student throughout the development process. A special event such as a presentation may provide a clearly summative assessment activity related to the portfolio. These assessments provide a broader, more in-depth look at what students know and can do than the traditional norm-referenced test or a written paper and pencil test. They contribute another view of a student's achievement beyond the traditional assessment provided through standardized tests.

A variety of models have been developed by states and others to test the portfolio assessment method. Some efforts were rigorously evaluated, providing data to improve this assessment method. Others have used it as an information assessment and communication tool. A few examples of these pioneer efforts are included.

The California Career-Technical Assessment Program (C-TAP) was designed by WestEd to examine various authentic assessment methods and their relationship to instructional improvement and outcomes. WestEd developed "The C-Tap Portfolio Guide to Evaluate Student Work," to identify requirements for the student portfolio and the criteria for assessment. They used this in the training sessions for teachers related to authentic assessment methods and how to score the portfolios. The California State Department of Education and the Sacramento County Office of Education were partners with WestEd in this study.

Oregon has legislated a "Certificate of Advanced Mastery" to document student achievement toward that state's academic and career related learning standards. The Science Portfolio is an optional part of the Golden State Examination, California Department of Education, 1994. Ohio developed a passport portfolio system which has been used as a model for other states (Arter, 1995).

Juneau, Alaska, has developed the Integrated Language Arts Portfolio for use in all primary grades. The portfolio replaces report cards and standardized tests, and provides information about the growth and achievement of each student. Not only one judgment by the teacher is included, but also work samples related to writing, speaking, listening, and reading (Arter, 1995).

Vermont uses the portfolio as a means for student to demonstrate to parents what has been learned in problem-solving and math competency (McLaughlin & Warren, 1991). This state has set comprehensive standards for what students will included in the portfolios. The portfolio is a communication device for students to share school progress with parents. As with many other portfolio performance assessment models, validity and reliability are still being clarified.

Portfolio assessments that are successful provide a clear vision of student skills to be addressed, student involvement in selecting what will be included in the portfolio, the use of a strong criteria for defining what is quality, and a plan for self-reflection by students in the development and evaluation of the portfolio. Portfolios offer a cumulative assessment method that provides a broad view of the student achievements related to specific standards.

Portfolios are useful in collecting materials and samples of projects which are illustrative of the student's work. Samples of work provide tangible results of work accomplished. The concern is with how these are viewed with more than a subjective criteria. Hymes recommends the use of portfolios with combinations of assessments to make a better informed judgment on the progress of students. Joan Boykoff Baron, Connecticut Mastery Test Program, said her state's rating system was based upon successful range finder techniques. Teachers are taught how to look for evidence of criteria within a range of criteria. In this way a more constant rating of portfolio work can be achieved. Vermont trains teachers to be assessors. They evaluate sample portfolios against a set of at least seven criteria. One criterion is that the information in the portfolio suggests that learning for a student was beyond basic knowledge levels and the required higher-order thinking skills to complete the project. The evaluator is taught how to develop the skill of determining degree to which student skills are above the basic knowledge used and what additional suggestions should be made to the student for constructive evaluation.

The C-TAP System developed by WestEd included a scoring rubric that assisted with one of the key concerns of authentic and performance assessment. The Dimensional Scoring system identifies the "dimensions" and then provides detailed information for each of four levels of scoring. Dimensions include: Content Knowledge, Applications of Content Knowledge and Skills, Career Preparation, Self-evaluation, and Written Communications. In addition, the Holistic Scoring System provides a method of examining the document as a whole with a defined criteria, including score ranges from advanced, proficient, basic, to limited. Each has a very clear set of criteria to define what that level represents. Detailed work sheets and examples provide a solid basis for portfolio assessment that is valid and reliable [WestEd (b), 1998].

The portfolio method provides one means of authentic assessment for the student to develop a cummulative document of achievements. The project is also a cumulative assessment and lends itself to not only documenting achievements, but also to demonstrate how to apply learning to an authentic context.

Summary. The portfolio method is a cumulative form of assessment, allowing for students to add materials to the folder throughout the program. The types of materials gathered may be based on a set criteria established by state or local districts, with student selection of the specific examples of achievement. The portfolio objectives and representative materials may both be selected by the student. As a low-stakes assessment type, these may serve both the student as an in-course monitoring of progress as well as a program improvement tool for the teacher. As a high-stakes assessment whereby decisions about the total success of the student in the program, it may not be as useful.

Project Assessment. A second cumulative assessment, the project assessment provides an in-depth, hands-on exploration of a topic, theme, idea, or activity resulting in a product, performance, or event for assessment (Katz & Chard, 1989). The goal of a project is to learn more about the topic rather than to seek "right answers" to questions posed by the teacher. A key element is that students, as well as teachers, may pose the question to be explored. The project should not be external to the learning environment, but rather the method used for learning.

Project work provides students with opportunities to apply skills, learn proficiencies, utilize intrinsic motivation, and determine what they will choose. Students are accepted as experts (Katz, 1994).

WestEd [WestEd (c), 1998] identifies key features as follows:

1. Hands-on applications of knowledge and skills in a purposeful, authentic activity
2. Integration of knowledge and skills even across several subject areas
3. Focus only on one or two content standards

Students need to explore a complex, yet realistic question, problem, or activity over time. As they become actively involved in creating products, performances, or events related to the question, they learn more than independent knowledge and skills.

Katz (1994) discusses the three phases which constitute the project assessment method:

1. Phase I---- Getting Started
2. Phase II--- Field Work
3. Phase III-- Culminating and Debriefing Events

In Phase I, Katz discusses selection of topics. Key points include: 1) select with students the topics closely related to their experiences, 2) create topics that can expand beyond a single content area to provide growth, 3) consider topics rich enough to be explored over time, and 4) consider topics that relate to school, work, or combinations. Phase II provides activities that allow for investigation, including field trips, events, and other activities. During this phase, development takes place, including a period for observation, construction of models, exploring, predicting, discussing, and recording (Chard, 1992). Phase III provides for a conclusion, presentations, and reporting results.

WestEd has developed criteria for guidance with project assessment just as it has with portfolio assessment. The structure suggested for project assessment contains four phases:

1. Planning and organizing the project
2. Researching and developing the project
3. Producing the final product
4. Presenting the final product

WestEd with the C-TAP project has designed guidelines for each phase. For example, the planning and organizing phase includes project idea and topic development, project purpose and

goals, the design of the general process for completing the project, identification of resources needed to complete the project, collection of evidence of progress, and establishment of time lines for completion.

The C-TAP Project Guide for Evaluating Student Work identifies the requirements for assessment for both the dimensional scores and holistic scoring. The requirements are similar to those discussed for the portfolio assessment. The requirements are specific and include descriptions of each phase to help answer various questions: What does the overall project need to be? What is included in the plan? What constitutes evidence of progress? This guide provides a rubric for both the student and the teacher to assist with the assessment as well as learning activities.

Project assessment provides the opportunity for students to apply their knowledge and skills to a long term complex set of problems and circumstances linked to reality. The concern over validity and reliability with the other authentic assessment methods is present, but to a lesser extent than with the other two types being discussed.

Summary. The project method is a cumulative assessment that provides a view of students performance over a sustained period of time, although it may not provide the opportunity to test all skills acquired by the student or all of those identified as part of the program objectives. It is one type of authentic assessment that when combined with others may provide part of the whole assessment picture.

Scenarios. A third type of authentic assessment is the scenario. Written scenario assessments depict complex and realistic problems that workers and individuals confront in a given context (NSSAC & WestEd, 1998). The context may be related to work, family, the community, or other real-world situations. Response to the scenario demonstrates the individual's ability to apply knowledge and skills to real world situations.

The Education Development Center (EDC) gives the following definition for a scenario. "A scenario presents a real-life work situation and includes a routine procedure and unanticipated problem the student must master" (Malyn-Smith & Leff, 1997). EDC developed scenarios for the Bioscience National Skill Standards, which they directed.

Scenarios offer the opportunity for students to use a combination of technical skills, workplace skills, related academic skills, problem-solving, creativity, and other higher-order thinking skills to solve the problem posed in the scenario assessment. The problems posed are from real-world situations in business and industry, communities, and other settings.

Scenarios are more successful when a combination of stakeholders develop them. Certainly, workers and supervisors from the specific occupational cluster and positions within the cluster being studied need to be a prominent part of the stakeholder group involved in the development.

Educators can contribute to the process by determining how particular scenarios would be useful for the classroom.

The Manufacturing Linkage Project, led by the state of Indiana and V-TECS, has also been involved in developing assessment scenarios with industry. The Manufacturing Linkage Project has developed scenarios for virtually all of the Manufacturing Linkage Core Duties and Tasks.

The Manufacturing Linkage Project, uses a computer based development system called "PROFS" to develop scenarios with a structured format. This format has been established to easily create a large data bank of scenarios. The defined format will allow for various states to develop additional scenarios that can be added to the data bank. The Manufacturing Linkage scenarios process uses a trained scenario developer to work with individuals in a specific industry work function to establish the problems to be presented.

Building on the experiences learned through the Manufacturing Linkage Project, V-TECS has developed a prototype format for potential implementation. The prototype is designed for instructional and assessment of student learning (Losh, 1998).

The V-TECS prototype scenario includes:

1. Occupational duties and tasks identification
2. Required academic skills identification
3. V-TECS workplace skills identification
4. Applicable National Standards
5. Conditions of performance
6. Summary of performance
7. Performance criteria
8. Student instructions
9. Teacher instructions
10. Assessment criteria
11. Resources utilized
12. Procedures

The prototype provides a guide for instructors and workers to develop scenarios that include like format and specifications for each component. The scenarios developed under this guide have a performance and assessment criteria established as part of the process, thus allowing for quality control is a strong element in this model.

The C-TAP project in California used scenario assessments based on criteria established by WestEd. The conclusion was that the many variables involved with scenarios created certain problems related to reliability. Scenarios structured around a smaller number of skills or standards were more definitive and were able to be assigned evaluation criteria without large numbers of variables. In scenarios that relate to a large number of skills or standards, the acceptable solutions multiply thus causing difficulty in scoring. WestEd worked with several of

the National Skill Standards Projects to use national standards as the base for developing the scenarios. Each of these action research projects has produced products that will assist in the design and development of assessment scenarios.

WestEd (1998) identifies a process that includes the following steps:

1. Involve all stakeholders
2. Identify and weigh importance of standards
3. Conduct scenario reviews
4. Conduct pilot tests
5. Refine items based on pilot feedback

The use of multiple groups of stakeholders provided scenarios that could be used in a variety of settings. Targeting certain key standards and ensuring their adaptability to scenarios is key to successful development. Industry and education wrote scenarios together to provide all aspects needed. Scenario reviews against criteria ensure that they are written in the same language, that all criteria are addressed, that they were free of political bias, and that they are realistic, relevant, and measurable. Pilot tests ensure validity. Refinements are needed to improve weak items.

The structures of C-TAP scenarios include five models:

1. Means-end
2. Crisis and follow-up
3. Roles and responsibilities
4. Developing recommendations
5. Competing clients or priorities problem situation

Means-end scenarios address general problem-solving strategies in a skill area related to a specific job. A hypothetical situation is established and the respondents are primed to use analysis processes to break the problem into parts to solve.

Crisis and follow-up scenarios provide a critical situation to be resolved. Respondents are asked to deal with the crises, but also to establish how they would develop long-term strategies to eliminate the problems.

Roles and responsibilities scenarios identify workplace situations in which respondents are requested to delineate each worker's roles and responsibilities. This involves understanding the larger scope of a business.

Developing recommendations scenarios present a problem that is to be analyzed and a list of recommendations to be developed for a client or customers. Justifications are requested.

Competing clients and priorities scenarios ask respondents to prioritize competing tasks and explain how to communicate to others about these priorities.

WestEd (1998) identifies the Means-end as the most used scenario type in the C-TAP project. It provided the most potential scenarios addressing the criteria established.

The Family and Consumer Sciences National Standards (V-TECS, 1998) included sample scenarios for two occupational areas. The development process took place after the standards, competencies, academic proficiencies, and process questions were established. The stakeholders included several levels of restaurant and hotel workers, supervisors, and managers, as well as educators. With a set of guidelines, standards, and a short period of training, the group was able to develop a large number of quality scenarios directly related to each standard. Processes for validation and reliability are still in progress. Each of the groups identified above have established criteria and structures for developing scenarios.

The Education Development Center (EDC) scenario format includes:

1. The problem or the question
2. The workplace setting
3. Key competency areas demonstrated
4. Tasks for routine procedures
5. Skills, knowledges, and attributes

EDC states that the developers need to be trained to work with industry to establish the scenarios. Questions they pose to the groups involved in the development are: 1) How will we know the problem has been solved? 2) What do we need to know and be able to do? 3) What resources are needed to complete the scenario? 4) What questions can be used to guide the scenario process?

Education Development Center (EDC) has developed scenarios for all of the National Skill Standards for Bioscience, and has provided direction for the development of the sample scenarios developed for the National Standards for Family and Consumer Sciences. The scenarios developed for the Bioscience Standards are very specific and have been validated. The scenarios were created for learning as well as evaluation purposes.

V-TECS currently is reviewing and updating the *Taxonomy for Essential and Related Academic Skills*. As part of the update, difficulty levels are being assigned to the skills. This will be particularly valuable for performance assessment purposes related to current vocational-technical education, school-to-work, tech-prep, workforce development, and academic programs (Losh, 1998).

Summary. Like other authentic assessment methods, scenarios are challenged with validity and reliability concerns, particularly when they are developed in one locality with a limited number of stakeholders.

WestEd and V-TECS in interviews, both reported a need to explore further use of written performance tests in conjunction with scenario assessments to ensure validity. The use of more than one set of measures to determine outcomes is not a new concept. If multiple types of assessment were to be used, a more valid and reliable set of data would be available to make judgments about student success and program improvement.

Scenarios, like projects and portfolios, add dimensions to assessment that traditional testing cannot. Further research is required to bring all the components of assessment together for both valid and reliable assessment results to be present.

Vocational-Technical Education Alternatives in Action

Over the years, assessment activities in vocational-technical and workforce related education programs have been promoted by several groups with a variety of objectives and activities. Some of these groups are directly related and funded by vocational-technical education funding, such as the vocational-technical education student organizations and V-TECS. Other groups may be non-profit or profit groups and may or may not receive support from states, such as ACT and NOCTI. Several of these provide resources related to testing and assessment for occupational, related academic, and employability skills. Others, such as the vocational-technical education student organizations provide performance assessment activities. A testing system and materials are provided by American College Testing (ACT), which established Work Keys. The Vocational-Technical Education Consortium of the States (V-TECS) provides supports through a system of standards and test banks for competency-based testing and assessment. The National Occupational Competency Testing Institute (NOCTI) conducts the Student Occupational Competency Achievement Testing (SOCAT) and National Occupational Competency Tests for individuals and materials to states that join their consortium.

Vocational-Technical Education Student Organizations. The student organizations include Future Business Leaders of America (FBLA), Distributive Education Clubs of America (DECA), FFA, representing Vocational Agricultural Programs, Future Homemakers of America (FHA-HERO), Health Occupations Schools of America (HOSA), Office Education of America (OEA), and Vocational Industrial Clubs of America (VICA). These organizations are dedicated to providing co-curricular activities at the secondary and to a limited extent at the postsecondary level in vocational education content areas. Each of these groups conducts a variety of activities including leadership, community support, and proficiency activities. Each of the student organizations conducts competitive and semi-competitive activities that provide the opportunity for individual and groups of students to exhibit their skills related to occupational skills, employability skills, academic skills, and team skills. Local, state, and national contests and related activities are held to assist students in measuring their skills against the criteria established by business and industry. Students not only take written examinations, but also perform tasks to meet the specifications required of a worker in a field. Examples of these performance assessments include such activities those related to VICA where students, individually or as a team, may be demonstrating how to lay a brick wall, repair hydraulic brakes

on a truck, or complete a CAD drawing to the specifications of the project. Students in HOSA may be given a problem to be solved in a hospital when an emergency situation arises. Students in FBLA might be asked to take an in-basket assignment to complete a project that includes use of the various computer programs, communicating with others, and producing a product. DECA students may be asked to develop a sales campaign for a new product. FHA-HERO may be involved in either occupational or family skills demonstration projects. A FHA- HERO student in food preparation may be asked to prepare a menu for a resort and develop one dish. A student participating in a course in family life education may be asked to resolve a family resource problem where competing needs of family members are at stake.

The events and the criteria for measurement for all areas and levels are developed by business, industry, and community representatives related to each area. National criteria is established in each area and adopted by the state and local level groups. Typically, students participate in local school level events to determine how they measure against the criteria. Those who are at a level of competency that the state determines is adequate may participate in the state events. Here typically, those in the first four or five places are selected to participate in the national competitive events. At the state level, two types of criteria are often applied. One is the measurement against a set criteria. The second is the competition amongst those in the event, as it might be in an application situation for a job. VICA also is involved in an international competition. Many nations send their champions to this biennial event. In this competition, world benchmarked skills and knowledge are used as the criteria.

Not only do students participate in occupational skill performance events, but often a written test is given prior to the performance activity to determine if the student possesses the necessary academic and occupational knowledge to perform the skill successfully. In addition, events related to Total Quality Management are offered at the state and national levels for teams of students. This have been carefully developed and validated by business, industry, and education. Employability skills and higher order thinking skills are woven into the occupational performance events as well as separate events.

The vocational-technical education student organizations provide a matrix of performance and authentic events that are valid, reliable, up-to-date, and sanctioned by all levels of business and industry. The events are revised on an on-going basis, based on the changing needs of business and industry. The criteria change in accordance with national and world standards for work related activities.

System for Test Item Banks. V-TECS is a consortium of 20 states with the goal of promoting competency-based vocational-technical education. During the early 1980s, they began to construct their own competency-based tests in response to a growing interest in better assessment and credentialling for vocational education students. The test banks are continually being expanded with the addition of new banks each year.

Four main products and services of V-TECS are:

1. Analytical tools, including catalogs of over 200 lists of duties and tasks for specific jobs,
2. Instructional tools, including enabling competencies and lists of related academic skills,
3. Assessment tools, including criterion-referenced test items and performance-based items for 35 job areas,
4. V-TECS DIRECT, a software package designed to store and retrieve V-TECS materials.

The V-TECS item banks have been developed by 11 states and the central office. These are developed according to the following V-TECS process:

1. Validating the task lists, performance objectives, and performance steps,
2. Writing test items to match the validated and updated task lists,
3. Reviewing and editing test items with a group of writers, test item construction experts, subject matter experts, and a sample of workers from the field,
4. Field testing item banks to ensure clear, reliable material free of gender and racial bias,
5. Editing and completing final item banks (McCage, 1993).

Border, B. (1993). *Education driven skill standards systems in the United States, Volume 2.* Washington, D.C.: Institute of Educational Leadership. Their work with test banks started in the early 1980s and has moved through many stages to the present item banks.

The Work Keys System. Work Keys is a system being developed by ACT for instruction and testing of general workplace competencies and employability skills. The system has four interrelated components for each of eight skills areas. The testing component includes 12 workplace skills tests or assessments of a criterion-referenced nature which require written response. The instruments will measure the level of academic competency demonstrated by the individual. Test formats include multiple choice, constructed response, and computer-adaptive. The job-profiling component includes a profile of competencies and academic skill levels required for an employee to perform selected jobs successfully. The third component of the system is instructional materials to assist the learner and the teacher in steps to improve and broaden the learner academic skills. The fourth component of the system is a record and reporting system, much like a transcript system available for students and educators to use. Students may have their records sent to perspective employers.

The assessments measure work-related basic academic skills with emphasis on workplace applications of those skills, and are criterion-referenced, with the examinee being evaluated against content attainment as opposed to other test takers. The six tests used as part of the assessment include:

1. Reading for Information
2. Listening and Writing
3. Applied Mathematics
4. Applied Technology
5. Teamwork
6. Locating Information

ACT intends to include tests on speaking, observation, motivation, learning, and managing resources (Wirt, 1994).

Testing Program. The performance-based testing program was created in the early 1970s, with the original mission to develop examinations to assess occupational competencies of vocational education teachers. NOCTI is the primary provider of vocational teacher competency exams in 60 specific areas. NOCTI does testing for industry with the Industrial Occupational Competency Test (IOCT). These tests are geared to pre-employment testing; in-house training needs, specific to the training needs of an employer; certification of skills; and promotion of employees.

The tests developed by NOCTI are both written and performance based. The written tests are criterion-referenced and measure the student or employee's ability in a specific occupational area. The performance tests require the student to perform various tasks. Evaluators must be from the industry. These tests are secure and held by NOCTI.

NOCTI has developed their tests using the DACUM process to determine task lists and performance criteria. Team members and subject matter specialists have designed the test items. These are then reviewed by testing experts. The items are field tested. Content validity is achieved by having workers develop the task lists and provide feedback on the test content. NOCTI calculates reliability for the written tests using the Kuder-Richardson method of internal consistency of the test items (NOCTI, 1994).

There are also other groups working toward types of tests and assessments in vocational-technical education, at present.

VI. Implications and Recommendations

A new era of continuous improvement for vocational-technical education is essential as education seeks to improve and the economic development system moves to correct the inadequacies of the past in the workforce. Business and industry need a technically skilled and knowledgeable workforce to provide the goods and services. Education needs to provide all stakeholders with the opportunity for input into the system, to understand the progress reports, and to move forward with defined objectives and criteria for assessment related to needs at various levels.

Authentic assessment, coupled with written performance assessment can provide the opportunity for vocational-technical education and related academics to provide clear, accurate, and comprehensive assessment data for all stakeholders. Together, these assessment methods provide rigor in quantitative data and added richness of qualitative data. This combination can be used to improve vocational-technical education and related academics programs along with student results related to the essential objectives of education and the needs of business and industry.

Overall, the following implications and recommendations are offered:

1. Rigorous formal development of performance and authentic assessments is needed to give valid and reliable qualitative data for vocational-technical educators at both state and local levels. Research in the past three years has provided the qualitative tools to develop formal models for use within vocational-technical education to assess technical skills, workplace skills, and related academics skills.

   Consideration may be given to the following:

   A. Use research findings from various centers and researchers as the base for establishing guidelines for the development of comprehensive systems.
   B. Develop a strategy for states to address the 1998 Carl D. Perkins Act and other workforce development acts, along with various state requirements, using both the qualitative and quantitative methods of assessment.
   C. Design technological tools to assist with the process for development, data collection, and interpretation.

2. Individual elements that could contribute to an integrated system for assessment and instruction have been explored, further refined, but as yet do not come together to provide a "triangulated" or multifaceted system with rigor. Certain components and connectors still need to be designed.

   Consideration may be given to the following:

   A. Leadership is needed to identify all parts of the system and provide models that will work for all areas of education as well as business and industry.
   B. Guidelines, in-service, and instructional materials are needed to assist states and local districts in establishing quality systems that will be recognized by all stakeholders.
   C. Components of the system need further research and clarification, through bringing together researchers, practitioners, and end users of the data.

3. Assuring validity and reliability in the assessment system is critical. Authentic assessment may need to be joined with written performance testing to bring together both the qualitative and quantitative elements needed to provide valid comprehensive data.

   Consideration may be given to the following:

   A. Future designs that provide a comprehensive examination of what an individual knows and can do, in relation to technical skills, workplace skills, and related academic skills, through using combinations of the qualitative and quantitative tools available.
      1) Build a system to include both written performance and authentic assessment situations that complement each other.
      2) Include rigor in the design process, implementation, and scoring phases.
      3) Establish an assessment design for large scale assessments, but with connections to those conducted in less formal settings. Provide mechanisms to collect rich data into the larger system of assessment.
      4) Develop a system that carefully examines the use of authentic and performance assessments in relation to written tests when high-stakes assessment is developed.
      5) Provide common language and data collection methods that will communicate between low-stakes and high-stakes assessment modes as well as between quantitative and qualitative assessments.

B. The following guidelines may be used to develop valid and reliable assessment systems:
   1) Identify key elements and procedures for both authentic and written performance tests.
   2) Provide procedures for using these tests, including considerations for various conditions, available materials and equipment, geographic location, environmental concerns, student assessability needs, and other anomalies.
   3) Provide scoring rubrics and procedures to ensure validity and reliability.
   4) Develop a training package for developers and users to ensure rigor.

4. The criteria for scoring authentic and performance assessments is an extremely important element of the assessment system for it to have rigor. A rubric for scoring is needed to accompany authentic and performance assessment modes. This appears to be the most frequent concern related to reliability of authentic assessments.

   Consideration may be given to the following:
   A. Design several prototypes for various authentic assessment modes for use by those who will score internally and by those who will score externally.
   B. Develop instructional guides and in-service programs for training those who will score both the low-stakes and high-stakes authentic assessments.
   C. Design an evaluation tool to assist those who will score to improve their abilities as they continue to rate authentic assessment.
   D. Design guidelines that incorporate "triangulation" or use of multiple assessment methods in the overall scoring of the quantitative and qualitative assessments into multiple or single ratings.

5. Begin the process. The dialogue related to assessment and testing has long been casual and less serious than it must be in the future for vocational-technical education and other related workforce development programs.

   Consideration may be given to the following:
   A. Formulate a design for action.
      1) Include those involved in vocational-technical education, workforce development, and academic education.
      2) Include researchers, issues, and systems designers.
      3) Develop a plan of action and invite input.
   B. Promote and encourage development of assessment systems, but carefully design the critical pieces.
   C. Develop a training program to ensure rigor, continuous improvement, and results.

References

American Education Research Association. (1985). American Psychological Association, and National Council on Measurement in Education Guidelines. Washington, DC: American Education Research Association.

American Youth Policy Forum. (1994). Building a system to connect school and employment. Washington, DC: American Youth Policy Forum.

Arter, J. (1995). Portfolios for assessment and instruction. Portland, OR: Northwest Regional Educational Laboratory.

Baker, E., & R. Linn. (1991). The center for research on evaluation, standards, and student testing monograph. Los Angles, CA: The Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E., O'Neil, H., & Linn, r. (1991). Policy and validity prospects for performance-based assessments. American Psychologist, 48, 1210-1218.

Border, B. (1993). Education driven skill standards systems in the United States, Volume 2. Washington, DC: Institute of Educational Leadership.

Border, B.A. (Ed.). (1998). National standards for family and consumer sciences. Decatur, GA: Vocational-Technical Education Consortium of States, Southern Association of Colleges and Schools.

Chapman, C. (1991). What have we learned from writing assessment that can be applied to performance assessment? Presentation at ECS/CDE. Alternative Assessment Conference Proceedings, Breckenridge, CO.

Chard, S. (1992). The project approach: a practical guide for teachers. Edmonton, Alberta: University of Alberta Printing Services.

Cuban, L. (1989). Enduring resiliency: enacting and implementing federal vocational education legislation, work, youth and schooling historical perspectives on vocational education. Stanford, CA: Stanford University Press.

Darling-Hammond, L. (1993). Authentic assessment in practice: a collection of portfolios, performance tasks, exhibitions, and documentation. National Center for Restructuring Education, Schools, and Teaching Monograph. New York, NY: Columbia University.

Darling-Hammond, L. (1994). Setting standards for students: the case for authentic assessment. Educational Forum, Fall, 59, 14-21.

Doolittle, P. (1994). Teacher portfolio assessment. ERIC Clearinghouse on Assessment and Evaluation. Washington, DC: U. S. Department of Education, ERIC/AE Digest.

Education Development Center (Leff, J., Ed, 1995). Gateway to the future: skill standards for the bioscience industry. Newton, MA: Education Development Center.

Elliott, S. (1995). Creating meaningful performance assessments. Reston, V.A.: Council for Exceptional Children.

Felstehausen, G. (1995). Authentic assessment for occupational competency for career and technology education. Final report. Year two. Lubbock, TX: Texas Tech University.

Fuchs, L. (1995). Connecting performance assessment to instruction: a comparison of behavioral assessment, mastery learning, curriculum-based measurement, and performance assessment. ERIC Digest E530. Washington, DC: U.S. Department of Education, ERIC.

Halperin, S. (Ed.). (1998). The forgotten half revisited. Washington, D.C.: American Youth Policy Forum.

Hayes, R. & Wheelwright, S. (1984). Restoring our competitive edge: competing through manufacturing. New York, N.: John Wiley.

Herman, Joan L. (1992). What research tells us about good assessment. Education Leadership 49(8), 74-77.

Herman, J. & Golan, D. (1991). Effects of standardized testing on teachers and learning--another look. Los Angeles, CA: Center for the Study of Evaluation.

Highlights, fourth annual report 1992-93. (1993). Canberra, Australia: National Training Board.

Hutchinson, MN. (1995). Performance assessments of career development. ERIC Digest. Ottawa, Ontario.: Canadian Guidance and Counseling Foundation.

Hymes, D.L., Chafin, A.E. & Gonder, P. (1991). Testing and assessment: problems and solutions. Arlington, VA: AASA.

Inger, M. (1993). Authentic assessment in secondary schools. Institute on Education and the Economy Brief, 6 Mar.. New York, NY: Columbia University.

Jacobs, J. A. (1992). Certification and accreditation law handbook. Washington, DC: American Society of Association Executives.

Jessep, G. (1991). Outcomes: NVQs and the emerging model of education and training. London, England.: The Falmer Press.

Jennings, J. (Ed.). (1993). National issues in education: the past is prologue. Bloomington, IN: Phi Delta Kappa.

Katz, L. (1994). The project approach. Eric Clearinghouse on Elementary and Early Childhood Education. Urbana, IL and Washington, DC: U.S. Department of Education, ERIC/CASS.

Katz, L. & S. Chard. (1989). Engaging children's minds: the project approach. Norwood, NJ: Ablex.

Lamm, T. (1995). Fairness in performance assessment. ERIC Clearinghouse on Counseling and Student Services. Greensboro. NC, and Washington, DC: U.S. Department of Education, ERIC/CASS.

Linn, R. (1993). Educational assessment: expanded expectations and challenge. CSE Technical Report 351, CA: CRESST.

Linn, R. & Baker, E. (1993). Comparing results from disparate assessments. The CRESST Line, CA: CRESST.

Linn, R.M., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment: expectations and validation criteria. Educational Researcher. 20(8): 15-21.

Losh, C. (1998). The linkage system: linking academic content standards and accepted skill standards monograph. Decatur, GA: Vocational-Technical Education Consortium of States, Southern Association of Colleges and Schools.

Losh, C. (1998). Virtual training and education scenario prototype (V-TECS)monograph. Decatur, GA: Vocational-Technical Education Consortium of States, Southern Association of Colleges and Schools.

Malyn-Smith, J. & Leff, J. (1997). Scenario learning. Newton, MA: Education Development Center.

McCage, R. & Hathaway, B. (1993). Essential competency based tools for preparation of the Workforce. Decatur, GA: Vocational-Technical Education Consortium of States, Southern Association of Schools.

McLaughlin, M. & S. Warren. (1994). Performance assessment and students with disabilities: usage in outcomes-based accountability systems. Reston, VA: The Council for Exceptional Children.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, , (2), 13-23.

Meyer, C.A. (1991). What's the difference between authentic and performance assessment'? Education Leadership 49(8), 39-40.

Munch, J. (1991). Vocational training in the federal republic of Germany Berlin: European Centre for the Development of Vocational Training.

National Center on Postsecondary Teaching, Learning, and Assessment. (1998). Creating a dynamic system of assessment.(Presentation), The NCTLA Assessment Institute, Burlington, VT.

National competency standards: policy and guidelines, second edition. (1992). Canberra City, Australia: National Training Board.

National Skill Standards and Assessment Collaborative & WestEd. (1998). Cross-industry assessment and certification, framework and implementation guide. San Francisco, CA: West Ed.

National Skill Standards Board. (1998). Annual report for 1997-1998, working for better jobs and better business results. Washington, DC: National Skill Standards Board.

National Training Board. (1992). Australian standards framework, NTB network. Canberra, Australia: National Training Board.

Nothdurft W. E. (1989). School works: reinventing public schools to create the workforce of the future. Washington, DC: The German Marshall Fund of the United States.

Nuttall, Desmond L. (1992). Performance assessment: the message from England. Education Leadership 49(8), 54-59.

O'Neil, H. F. (Ed.). (1997). Workforce readiness, competencies and assessment. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Oregon Department of Education. (Jonnerg, 1997). Oregon standards. Salem, OR: Oregon Department of Education.

Pandey, T. (1990). Authentic mathematics assessment. ERIC clearinghouse on tests, measurement, and evaluation. Washington, DC: U. S. Department of Education, ERIC/CASS.

Peretz, J. (1992). Skills standards and certification in the U.K. First draft of a paper prepared for the Commission on Work Based Learning. London, England.

Pipho, C. (1988). Testing debate. Paper presented at the Education. Commission of the States Conference.

Popham, W.J. (1993). Circumventing the high costs of authentic assessment. Phi Delta Kappan, Feb. 74(6), 470-73.

Pucel, D.J., Quetler, T.J., Damme, S., Warner, D. (1992). Performance based Occupational Affective Behavior Analysis (OABA) St. Paul MN: University of Minnesota.

Resnick, L.B. and Klopfer, L.E. (1989). Toward the thinking curriculum, Toward Thinking Curriculum: Current Cognitive Research. Alexandria, VA: ASCD.

Roeber, E. (1995) Guidelines for the management of performance assessments in large-scale assessment programs. Oak Brook, IL: North Central Regional Educational Laboratories.

Rosenbaum, J.D. & Kariya, T. (1989). Postsecondary education and pathways to work in Japan. Evanston, IL: Northwestern University.

Senge, P.M, Kleiner, A., Roberts, C., Ross, R., Smith, B. (1994). The fifth discipline field book. New York, NY: Doubleday.

Simon, Karen, &S. Gregg. (1993). Alternative assessment--can real-world skills be tested? Policy Briefs. Charleston, WVA: Appalachia Educational Laboratory.

Smith, S. & Moorhouse, M. (1994). Skill standards projects: profile report. (U.S. Department of Labor No. K-4316-3-00-80-31 and 0912-1-010197). Washington, DC: Aguirre International and CAL, Inc.

Stiggens, R. (1995). Sound performance assessments in the guidance context. ERIC Digest. Washington, DC: U.S. Department of Education, Office of Education Research and Improvement.

Stiggens, R. (1988). The design and development of performance assessments. Educational Measurement: Issues and Practice, 6(3), 33-42.

Student job ready assessments. (1994). Big Rapids, MI: National Occupational Competency Testing Institute.

Tannenbaum, J. (1996). Practical ideas on alternative assessment for ESL students. ERIC Clearinghouse on Languages and Linguistics. Washington, DC: U.S. Department of Education, ERIC.

Taylor, C. (1994). Assessment for measurement or standards. American Educational Research Journal, 31, 231-262.

Thurow. L. (1992). Head to head. New York: Warner.

WestEd. (1998). Developing a standards-based assessment system. San Francisco, CA: WestEd..

WestEd. (1998). The C-TAP portfolio guide to evaluating student work. Sacramento, CA: California Department of Education and Sacramento County Office of Education.

WestEd. (1998). The C-TAP project guide to evaluating student work. Sacramento, CA: California Department of Education and Sacramento County Office of Education.

WestEd. (1998). The C-TAP written scenario guide to evaluating student work. Sacramento, CA: California Department of Education and Sacramento County Office of Education.

Wiggens, G. (1989). Teaching to the authentic test. Educational Leadership, April 41-47.

Wills, J. (1997). Standards: making them useful and workable for the education enterprise. Washington, DC: Office of Vocational and Adult Education, U.S. Department of Education.

Wills, J. (1993). An overview of skill standards systems in education & industry (Vol. 1). Washington, DC: Institute of Educational Leadership.

Wills, J. (1993). Industry driven skill standards systems in the United States (Vol. 3). Washington, DC: Institute of Educational Leadership.

Wills, J. & Sheets, R. (1993). Skill standards systems in selected countries (Vol. 4). Washington, DC: Institute of Educational Leadership.

Wirt, J., Eisner, E. & Williams, C. M. (1994). Testing and assessment in vocational education (U.S. Congress, Office of Technology Assessment OTA-BP-SET-123). Washington, DC: U.S. Government Printing Office.